PB-217 633

STATISTICAL CONSIDERATIONS RELATED TO THE PLANNING AND OPERATION OF A NATIONAL AIR SAMPLING NETWORK

Eugene K. Harris, et al

Robert A. Taft Sanitary Engineering Center
Cincinnati, Ohio

May 1956

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

PB 217 633

STATISTICAL CONSIDERATIONS RELATED TO THE PLANNING AND OPERATION OF A NATIONAL AIR SAMPLING NETWORK

Eugene K. Harris and Elbert C. Tabor

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

ROBERT A. TAFT SANITARY ENGINEERING CENTER
Public Health Service
Cincinnati, Ohio

May 1956

In: Proc. 49th Annual Meeting APCA, Buffalo, N. Y., May 20-24, 1956, pp. 35-1 thru 35-18.

Statistical Considerations Related to the Planning and Operation of a National Air Sampling Network

Eugene K. Harris[1] and Elbert C. Tabor[2]

The national air sampling program supervised by the Robert A. Taft Sanitary Engineering Center, a research facility of the Public Health Service, United States Department of Health, Education, and Welfare, involves at present about 90 high-volume sampling stations and over 30 metropolitan areas.

Operations began in October, 1953, under a contract with the Army Chemical Corps to measure the concentrations of certain particulate constituents of the atmosphere over a number of American cities. From the start, the network has consisted of communities definitely interested in air sampling or at least willing to assist the Sanitary Engineering Center in carrying out such a program. An attempt has been made to obtain as wide a geographical distribution of sampling sites as possible. At least two sites have been chosen in each community: one or more in business or residential sections, another in a rural or suburban environment.

As a rule, at least one sample per week has been obtained from each site. Some stations have been sampling on an almost daily schedule. Each sample represents the total of particulate

---

[1] Analytical Statistician
[2] Acting Chief, Chemical Research and Development Section, Community Air Pollution Program
both of the Robert A. Taft Sanitary Engineering Center, Public Health Service, Cincinnati, Ohio.

Reproduced from best available copy.

matter, in sizes down to 0.1 micron, removed from approximately 2,000 cubic meters of air during a 24-hour period. Further details on analytical procedures as well as a comprehensive discussion of findings to date, with respect to particulate loadings and organic material, are given by Chambers et al. (1,2). In planning future operations of the program, attention has been directed to statistical considerations involved in its most effective use. The present paper will discuss some of these.

This air sampling program has been referred to as a "network" and, since this is a convenient term, it will probably be continued. It is not the best word for the purpose, however. Statistically, a network implies a group of connected points, each contributing to a common goal, such that the loss of even one point diminishes the likelihood of attaining that goal and the loss of more than some critical number of points makes the goal completely unattainable. For example, whenever these sampling sites are used to map fall-out from a single source of emission, they do constitute a network.

Except for such occasional joint action, we are dealing essentially with independent metropolitan areas, each producing its own air pollution pattern under the influence of chiefly local meteorological and topographical conditions. To be sure, two cities may show similar pollution patterns, but this usually connotes a similarity of local emitting and dispersing situations, rather than the effect of a common set of conditions.

Assuming independent areas, and limited analytical facilities restricting the extent to which they may be sampled over time and space, what are the possible objectives of this program, and how can statistical theory contribute to its planning and operation to satisfy those objectives? The objectives considered below are chiefly of a statistical nature. There are also important training and educational goals which will not be discussed here.

OBJECTIVES AND SAMPLING SCHEDULES

One primary object is to estimate trends in the concentrations of various air contaminants over a specific area. Another important objective is the detection of unusually high levels of pollution. There is serious question whether a single sampler can satisfy both these purposes. The former is served best by continuous low-volume sampling, while the latter calls for high-volume sampling at frequent intervals for short periods (eight hours perhaps). Sampling for 24 hours once every week or two attempts to serve both objectives but may, at worst, satisfy neither. It does not impose as great a load on the analytical chemist as would a program involving two samplers at each site.

This problem of one or two types of sampling schedules has not yet been resolved. One possible solution would be to

undertake only continuous, low-volume sampling in non-urban areas, both continuous low-volume and frequent short-period high volume sampling in certain metropolitan complexes, and, finally, intermittent high-volume 24-hour sampling in the remaining urban areas. In any case, intermittent sampling is going to play a major part in the program, if not, as heretofore, the entire role. This method of sampling raises some fundamental statistical considerations.

SAMPLING FOR TRENDS

Considered in terms of 24-hour samples for convenience, the chief problem is to assure a random selection from the infinite population of 24-hour samples possible during any extended period. No statistical analysis of sampling results is worth much unless it leads to generalizations regarding these populations, such as, for example, that the mean weight of acetone-soluble materials per cubic meter of air over Site X is not significantly greater in winter than in summer. The validity of such generalizations depends entirely on randomness in the data-collecting mechanism.

Specifically, a sample selection plan, applied to an infinite population of values, is defined as random if, on the average, the distribution of sample values reflects the distribution of values in the population. That is, the relative frequency of sample values within a certain range should equal, on the average, the true probability, defined as the area above that range under the probability distribution curve of the population.

This criterion of randomness would be satisfied if we were as likely to sample during any one 24-hour period as another. A completely random sampling design would call for an automatic sampler set to collect some specified number of samples at random intervals of time (e.g., at intervals ranging from 48 hours to 2 weeks between samplings).

An entirely automatic random sampler may not be realistic at present. A more practical plan, just as satisfactory from the statistical viewpoint, would require some central agency like the Sanitary Engineering Center to prepare and distribute a random sampling schedule for each area. An inexpensive timer could be attached to each sampler so that if, for example, the schedule called for a given sampling period to begin at 2 A.M., the operator would merely set the timer during his normal working day and then check on the following day to be sure the sampler had begun operating during the night.

Completely random sampling would not be ideal if certain days of the week were known to show consistently higher levels of air pollution than other days. The days would then group themselves into two or more groups, or strata, calling for stratified random sampling, whereby each group would be sampled in proportion to the number of days in the group. Within groups, sampling would be completely random.

Any benefit derived from stratified as against completely random sampling is measured in terms of the relative precision

with which averages are estimated. This precision, in turn, is measured inversely by the standard error of the mean.

In air sampling, two obvious possibilities for stratification are weekdays and weekends. Suppose θ_1, and θ_2 denote, respectively, weekday and weekend daily averages of some pollutant. We assume for simplicity that variation in weekday pollution, over any given period is the same as that in weekend pollution, and denote the common variance by σ^2. This assumption has been upheld in all the data analyzed to date. Then, if $\sigma_{\bar{x}u}$ represents the standard error of an average concentration obtained from completely random sampling and $\sigma_{\bar{x}s}$ the standard error of an average obtained by drawing 5/7 x 100 per cent weekday samples and 2/7 x 100 per cent weekend samples, statistical sampling theory shows that the ratio:

$$\frac{\sigma_{\bar{x}u}}{\sigma_{\bar{x}s}} = R = \sqrt{1 + \frac{10(\theta_1 - \theta_2)^2}{4\sigma^2}} \qquad (1)$$

The larger the difference between weekday and weekend average concentrations relative to the variation within each of these groups of days, the larger will be the ratio R and the greater the gain in precision afforded by stratified sampling.

This equation may be put in a form more suitable for discussion. Let $\theta_1/\theta_2 = k$, the concentration ratio (i.e., the

weekday average is k times the weekend average). Further, put σ^2 in terms of relative error; that is, write $\sigma^2 = c^2 \theta^2$, where θ is the overall daily average and c is the coefficient of variation, written as a decimal rather than the usual percentage. Then, since

$$\theta = 1/7\ (5k\theta_2 + 2\theta_2) = \frac{5k + 2}{7}\ \theta_2 ,$$

we may write

$$R = \sqrt{1 + \frac{10\ (k-1)^2}{c^2\ (5k + 2)^2}} .$$

Suppose, for example, that variation in the daily pollution of weekdays or weekends over a season is 75 per cent of the overall average pollution, i.e., $c = 0.75$. Then, if weekday pollution averages twice as high as weekend daily pollution ($k = 2$), R will equal $\sqrt{1.123}$, or 1.06. In other words, stratified random sampling will yield only 6 per cent greater precision than completely random sampling. Even if weekday pollution averages four times higher, stratified sampling affords only 15 per cent more precision.

Suppose that percentage variation is less; $c = 0.5$, say. Then if weekday pollution averages twice weekend daily pollution, stratified sampling offers 13 per cent greater precision. Again, if variation within strata drops to 33-1/3 per cent, under a weekday/weekend ratio of 2, stratified sampling offers 28 per cent greater precision.

Now we may examine some actual data relevant to the network. Table I lists average particulate loadings and weights of benzene-soluble organic extracts, with their standard deviations, taken from high-volume filters in downtown City P. from September, 1955, through January, 1956.

(Insert Table I)

Weekday and weekend samples show about the same variability, as assumed in equation (1). The average standard deviation is about 45 per cent of the overall mean for total particulates, 57 per cent for organic material. Clearly, neither of these measurements of pollution shows sufficient variation from weekday to weekend to warrant stratified sampling. Similar conclusions have been reached in other cities, where the national program has provided daily records for limited periods of time. Examination of daily AISI tape sampler data in City C. during the months of most intense staining, September through December (1952-54), has revealed only a slight average drop on Sundays and none on Saturdays. In fact, wherever data relevant to the network have been studied with respect to daily variations in air pollution, completely random sampling appears to offer practically as high precision as a more complicated random design.

This result does not in any way imply that the difference between average weekday and average weekend pollution is not statistically significant. A test of this difference would involve

Reproduced from best available copy.

the standard errors of the two averages, whereas equation (1), used to judge the case for stratified sampling, involves the standard deviation of a single day's pollution. If the average weight of total weekday particulates in Table 1 had been 238 gm. instead of 227, it would have been significantly higher than the average weekend value at the 1 per cent level of significance. Yet this would not have strengthened the argument for stratified sampling, because the ratio of the averages, k, would still have been much less than 2. The question with respect to choosing a random sampling design is not whether the average difference between weekend and weekday is statistically significant, but whether it is so great that, unless formally recognized through stratified sampling, precision in the estimate of the overall average will suffer materially.

SAMPLING FOR PEAK LOADS

Another basic statistical problem concerns the frequency of sampling. This is particularly important in satisfying the second-mentioned objective of the national air sampling program, the detection of unusually high levels of pollution. We would like to estimate, say, the 95-th percentile value of some air contaminant, i.e., the concentration exceeded in only 5% of all possible 24-hour (or, in general, X-hour) samples during a given period. Information of this sort would be useful in selecting areas for epidemiological or vital statistics studies to measure possible health effects of community air pollution.

Consider that we are planning to collect a series of n samples at a given site according to some random sampling design. How large should n be in order that some high percentile value in the underlying population may be estimated with reasonable confidence? Let m_p denote the (100p)-th percentile value of a given variable (e.g., sulfur dioxide concentration). If, as would be natural in the planning stages, we are unwilling to assume a particular mathematical form for the distribution of values in the population, we must resort to nonparametric (or distribution-free) methods of estimation. A formula useful to this purpose (Mood, 3) offers an upper confidence limit to the percentile m_p.

If x_r is the r-th value when the n observations are arrayed from smallest to largest, then the probability that m_p is less than x_r is given by

$$\Pr(m_p < x_r) = \sum_{i=0}^{r-1} \binom{n}{p} p^i (1-p)^{n-i}$$

Suppose we are interested in the 95-th percentile (i.e., $p = .95$) and are planning to collect 49 samples during a year. This number was chosen because it happens to be the largest value of n handled in the binomial tables at our dispoal and is not an unreasonable number. Using tables of the binomial distribution, one may check that $\Pr(m_{.95} < x_{49}) = .92$. In other words, even if we choose the maximum of a series of 49 observations we are still only 92% confident that the 95-th percentile value in the

population of samples is below this figure. If we choose the second highest, x_{48}, our degree of confidence drops to 71%.

If we relax our demands and set p = .90, then the maximum of 49 observations provides an upper limit with 99.4% confidence; the second highest an upper 96.3% confidence limit; the third highest, an upper 86.0% confidence limit. Thus, in the absence of any information regarding the form of the underlying distribution curve, even once-a-week sampling, does not give highly reliable estimation of annual peaks beyond the upper 10% point. To obtain very dependable information about high-level values during a given season would seem to require several years of such sampling under the same environmental conditions.

These results are useful in planning a sampling schedule because they are not based on any a priori assumptions regarding the distribution of measurements and are therefore conservative -- disclosing the minimum that can be expected from a given sampling frequency.

However, after a year of random sampling, we would have learned something of the distribution of measurements, and would be in a position to apply more powerful methods of estimating high-level ranges. As a rule, measurements of air pollution tend to be skewed with a long tail of high values, since observations must be non-negative. In such cases, a square root or logarithmic transformation may produce near normality, and the estimation of

an upper bound to, say, the 95-th percentile follows from existing theory on tolerance limits for a normal distribution (e.g., Wallis, 4).

As an example, Figure 1 shows a graph of atmospheric lead concentrations observed during a year of 24-hour sampling in downtown City C., plotted in geometric intervals on normal probability paper. The approximately linear plot indicates a log-normal distribution of lead densities. In Table II the mean and standard deviation of the logarithms are listed and Wallis' formula used to determine the upper 90 per cent confidence limit to the 95-th percentile lead concentration. This limit turns out to be 11.5 γgms./M^3. That is, we may be 90 per cent confident that, under the environmental conditions prevailing in downtown City C. during the year in which these samples were taken, no more than 5 per cent of integrated 24-hour lead concentrations lay above 11.5 γgms./M^3.

(Insert Figure 1 and Table II)

It is well to remind ourselves that peak values of integrated samples may be far from the maximal air pollution levels to which the community's population is being exposed. For instance, in the Leicester survey of atmospheric pollution (5), the highest hourly mean smoke concentration at the center of the city was almost double the highest daily concentration in summer and 23% greater in winter.

We have discussed two primary objectives of the "network", the detection of trends and of high-level concentrations. These objectives may be attained with no more than one sampling station in each metropolitan area. In addition, it might be thought desirable, within the limits of facilities for chemical analysis, to gather information concerning the contours of pollution over a given area. Here we would need a true network composed of stations suitably placed with respect to prevailing meteorological conditions and distance from sources.

From three to five stations have been established recently in each of several cities as part of the national program; however, samples at sites other than the downtown station have rarely been collected more often than once a week. Even under continuous sampling, five stations in a metropolitan complex would be barely sufficient for determining contours or gradients of pollution.

In view of limited facilities for chemical and physical analysis of samples, plus the desire on the part of many local governmental agencies to participate in the program, it does not seem wise to continue multistation sampling within any metropolitan area, but rather to maintain no more than two stations: one in the heart of the city, the other a control in the nonurban surrounding area.

On this basis, we classify the National Air Sampling Program as an "extensive" rather than an "intensive" study. Nevertheless, the sampling problems of any community cannot be fully

understood except as a result of intensive study of that community. The national program established to determine trends and maximal ranges for numerous contaminants in many cities, can help any one city only as a screening mechanism. It can alert a community to the existence of contaminants at potentially hazardous peak levels, or to the presence of peculiar local pollution characteristics that should have closer study. Conversely, it can indicate that a given community has no need to investigate certain air pollutants that may be of considerable concern elsewhere.

Finally, then, those in charge of the national program are faced with the problem of deciding how long to continue sampling over a given area before drawing conclusions relative to the importance of any contaminant in the atmosphere of that area. From the foregoing statistical considerations, and plain common sense, a minimum of two years' sampling on a schedule calling for no less than 26 high-volume sampler per year appears necessary. Continuous, low-volume sampling, if and where initiated, would generate at least another 26 samples per year at that sampling site. As far as practical, the sampling plan should be tailored to the area, taking advantage of whatever is known concerning variations in air pollution, in order to achieve as reliable a set of random samples as foresight will allow.

References

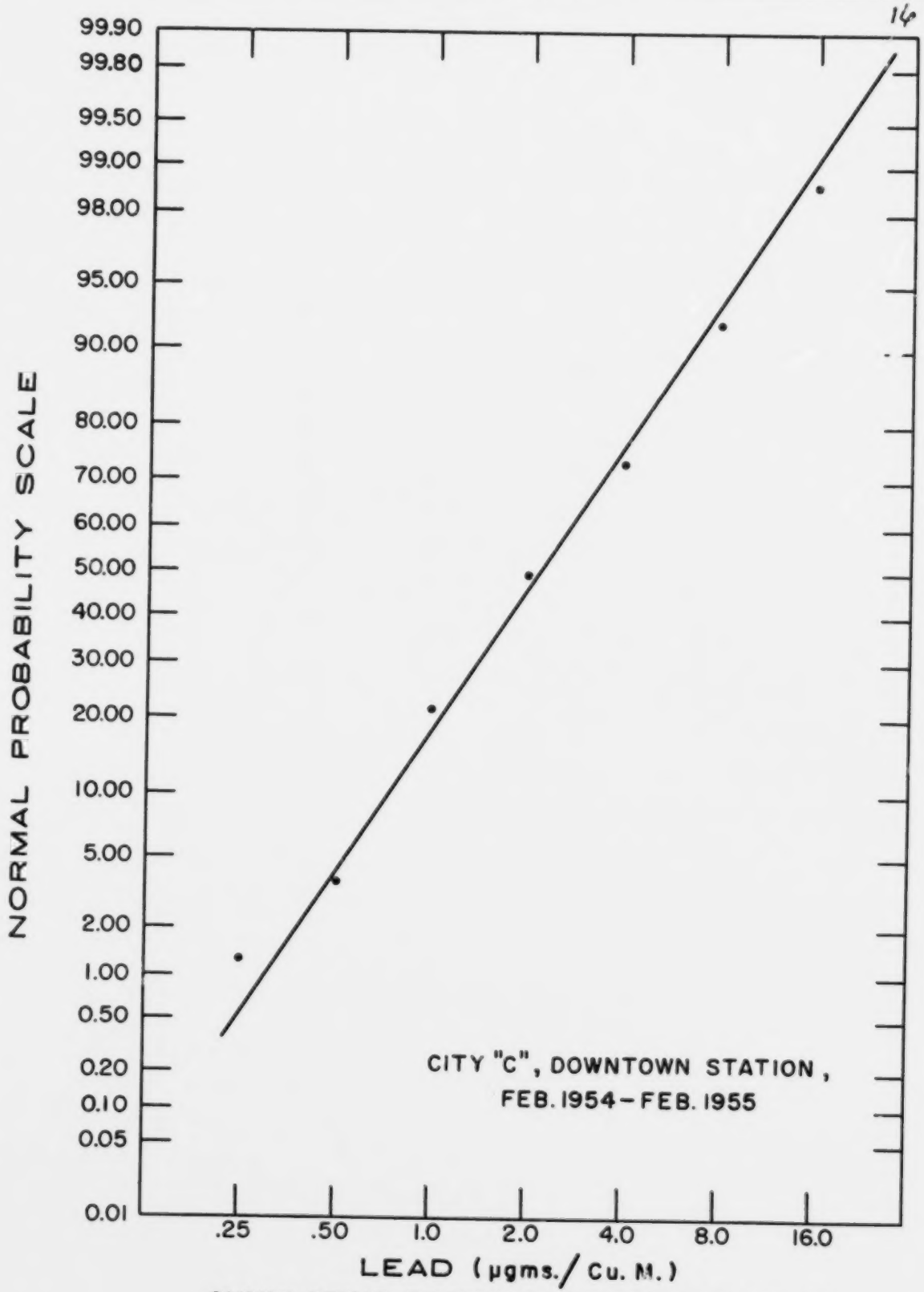
1. Chambers, L.A., Foter, M.J., and Cholak, J.: A Comparison of Particulate Loadings in the Atmospheres of Certain American Cities. Proceedings of the Third National Air Pollution Symposium, April 18 - 20, 1955, Stanford Research Institute, Stanford, California. Pp. 24 - 32.

2. Chambers, L.A., Tabor, E.C., and Foter, M.J.: The Characteristics and Distributions of Organic Substances in the Air of Some American Cities. APCA Annual Meeting Paper No. 55-12, Detroit, Michigan, May 22 - 26, 1955.

3. Mood, A.F.: Introduction the Theory of Statistics. McGraw-Hill Book Company, Inc., New York, 1950. P. 389.

4. Wallis, W.A.: Use of Variables in Acceptance Inspection for Percent Defective; Chapter 1, Techniques of Statistical Analysis by Eisenhart, Hastay, and Wallis; McGraw-Hill Book Company, Inc., New York, 1947. Pp. 7-20, esp. Pp. 19-20.

5. Atmospheric Pollution in Leicester; A Scientific Survey. Department of Scientific and Industrial Research, Atmospheric Pollution Research Technical Paper No. 1, H. M. Stationery Office, London, 1945.



CUMULATIVE FREQUENCY DISTRIBUTION OF ATMOSPHERIC LEAD DETERMINED FROM SPECTROGRAPHIC ANALYSIS OF 24-HR. SAMPLES

FIGURE 1

Table I

Averages and Standard Deviations of Total Particulate Loadings and Weights of Benzene Soluble Organic Material (gms./Cu.M.), in Downtown City P.: September 1955 through January 1956, by Weekday (Monday - Friday) and Weekend

Total Particulates

	Avg. and Std. Deviation	No. of Observations
Weekday	227 ± 99	94
Weekend	197 ± 101	35
Overall	219 ± 100	129

Organic Material

	Avg. and Std. Deviation		No. of Observations
Weekday	15.1	8.2	93
Weekend	12.3	7.5	35
Overall	14.3	8.1	128

Table II

Calculation of Upper Confidence Limit to 95-th Percentile of 24-Hour Total Atmospheric Lead at Downtown City C., Based on 78 Samples Obtained February 1954 through February 1955 (see Fig. 1) (Formula below given by Wallis (4))

Mean Log μ gms./M^3 = .3462 $\quad N = 78$

Std. Dev. Log Wt. = .378

Consider confidence coefficient γ = .90

Let K_γ = normal deviate exceeded with probability γ

= -1.2816 for γ = .90

K_{1-p} = normal deviate exceeded with probability 1-p, for p-th percentile

= 1.6425 for p = .95

Compute $a = 1 - \frac{K^2_\gamma}{2(N-1)} = .9893$

$b = K^2_{1-p} - \frac{K^2_\gamma}{N} = 2.6846$

$k_{p,\gamma} = \frac{K_{1-p} \pm \sqrt{K^2_{1-p} - ab}}{a} = \frac{1.6449 \pm \sqrt{.0498}}{.9893} = 1.888$

Then, upper limit with 90% confidence for 95-th percentile of log lead weight

= mean log $\pm$ $k_{.95,.90}$ $\times$ Std. Dev. Log

= .3462 $\pm$ 1.888 (.378)

= 1.0599

Antilog = 11.5 μ gms. Lead/M^3